Remarks

Claims 1-12 are pending in the application. Claims 1-5 and 7-12 are rejected. Claim 6 is objected to. Claims 1, 4, 5, and 10 are amended to correct informalities. The Specification is amended to correct clerical errors. No new subject matter is added. All rejections and objections are respectfully traversed.

Drawings

The Examiner objected to the drawings. A proposed drawing amendment to Figure 1 is submitted herewith. An improper duplicate use of reference number "100" is corrected. The blocks are amended to include descriptive labels as well as numeric labels.

However, the applicants respectfully assert that the Examiner's objection to Figure 5 under 37 CFR 1.84(p)(4) is incorrect. The Examiner is directed to MPEP 608.02(p)(4) which states "(4) The same part of an invention appearing in more than one view of the drawing must ***always be designated by the same reference character***, and the same reference character must never be used to designate different parts." (*emphasis added*)

Item 599 is a routine that determines whether a single volume remains. Therefore, the "599" designation is appropriate for both locations in Figure 5, because the same part of the invention is designation by the same reference character.

The Examiner objects to the Specification at page 9, line 22 – page 10, line 5 as failing to describe how a volume P_i is assembled. The Examiner is respectfully directed to the section titled "Assembling Volumes," beginning at page 10, line 7, i.e.,

"volume P_i is assembled, according to $m_i = \arg\min \nabla V(x, y, t)_{r,g,b}$,

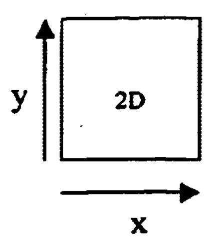
$S = V - \bigcup_{j=1}^{i} P_i$, until all the pixels of the volume are removed from the set S,"

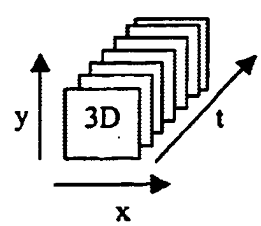
and following paragraphs.

Claims 1, 7, and 12 are rejected under 35 U.S.C. 103(a) as being unpatentable over Wu (U.S. Patent 6,529,202 – "Wu") in view of Frisken et al., (U.S. Patent 6,603,484 – "Frisken").

First, it is well known to anyone of ordinary skill in basic geometry that an image can only have two-dimensions, and no more. Thus a ***region*** of an ***image*** also has two-dimensional (x, y) coordinates, as in Wu. Second, it is also known that a ***volume*** of a ***video*** must have three-dimensional (x, y, t) coordinates, where t is a temporal coordinate because videos are a temporally ordered sequence of frames. Typically, the temporal dimension is represented along the z-axis. A person of ordinary skill in the art would readily understand this distinction. A diagram is included below to help the Examiner distinguish 2D images in Wu from 3D volumes according to the invention.

Wu





Invention

The invention operates on a video, i.e., a sequence of frames. The invention must operate on frames of a video, because the invention assembles video ***volumes***, which are three-dimensional. To assemble the volumes, a feature vector is assigned to each pixel of the video, i.e., each pixel of each frame of the video, and selected pixels of the video are identified as marker pixels. Each marker pixel and pixels adjacent to the marker pixel are assembled into a corresponding volume if the distance between the feature vector of the marker pixel and the feature vector of the adjacent pixels is less than a first predetermined threshold.

It is very important that the Examiner understands that when growing a 3D volume, pixels adjacent to the marker pixel must include pixels in ***adjacent frames*** of the video. If only adjacent pixels of the same frame are assembled, then only a 2D ***region*** can be grown, i.e., two-dimensions.

Wu operates only on an image, which is two-dimensional. The reference to column 4 by the Examiner describes growing regions in an image by adding spatially neighboring pixels to the two dimensional region. Claimed is assigning a feature vector to each pixel ***of the video***. The invention grows volumes, which require a third dimension, i.e., a temporal sequence of frames. Wu only operates on pixels of an image, not all pixels in a video as claimed. Therefore, Wu can never be used to make the invention obvious.

The invention identifies selected pixels of the 3D video as marker pixels, more specifically the selected pixels have a minimum gradient. Wu, at col. 3, lines 66-67, uses *arbitrary* pixel of a single 2D image as the starting point of a two-dimensional

color region. Wu never operates on a video. Wu's arbitrary pixel in his 2D image can never make obvious selected marker pixels throughout a sequence of frames of a video.

Each marker pixel, which Wu does not teach, and pixels adjacent to the marker pixel are arranged into a corresponding *volume* if the distance between the feature vector of the marker pixel and the feature vector of the adjacent pixels is less than a first predetermined threshold. The adjacent pixels include pixels in adjacent frames. This should now be understood.

Increasing color regions in a 2D image as cited by the Examiner can never make obvious three-dimensional volume growing.

Wu can never assign a score to a volume, sort volumes, compare descriptors of volumes, or combine volumes because Wu never generates volumes.

Frisken generates and operates on detail-directed hierarchical distance fields (HDFs). The input to Frisken is an object represented by a graphics model. For example, the input to Frisken can be a parametric surface model, which define the surfaces of an object as a collection of primitives such as polygons, spline patches, or subdivision surfaces that are defined parametrically; or implicit surfaces which are represented by an implicit function; or sampled volume data which represent objects and models in an array of sampled intensity values on a regular or irregular grid; or a distance field which is a scalar field that represents the "distance" to a surface of an object for each point in the field.

None of the input to Frisken is a video. Figure 14 of Frisken shows cells of an HDF. Cell data can include sampled values for M sample points associated with a cell which include sample values for each sample point. The sample values for each sample point can include attributes such as a distance field vector, appearance parameters, and computational parameters. See col. 10, line 53 – col. 11, line 11.

The Examiner's assertion that each cell represents a video object is incorrect. Frisken never describes anything having to do with videos. In fact, each cell represents samples of part of a *model* representing an object.

The root cell (node) represents the bounding box enclosing a portion of the distance field representing object, see col. 9, lines 5-21. The Examiner's assertion that "the root node represents the entire video" is incorrect. A video is not a model representing the shape of an object. Frisken is irrelevant to what is claimed.

Claim 2 is rejected under 35 U.S.C. 103(a) as being unpatentable over Wu and Frisken in further view of Kim, et al., (U.S. Patent 5,502,510- "Kim").

Kim describes a temporal prefilter using a motion adaptive spatial filter for use in encoding a video signal. Kim filters a video to be encoded. Wu grow color regions in 2D images. Frisken represents models as HDFs. None of the references can be combined. None of the references operates on the same input, or even similar input. Nor are the outputs similar.

In claim 2, each pixel has spatial (x,y) and a third time (t) coordinates to indicate a location of the pixel and the volumes in a spatial-temporal collocated overlapping

scene of the video. At col. 3, lines 1-3, Kim describes "A video signal may be treated in terms of its 3dimensional, i.e., horizontal, vertical and temporal components; and described as a continuous function f.sub.3 (x,y,t)." This is well known, as discussed above. However, the Examiner ignores the explicit limitations recited in the claim, such as location of the pixel and the volumes in a spatial-temporal collocated overlapping scene of the video. The applicants request the Examiner reconsider and withdraw the rejection.

Claim 3 is rejected under 35 U.S.C. 103(a) as being unpatentable over Wu, Frisken, and Kim, in further view of Kumar, et al. (PUB. No. US 2001/0038718 "Kumar").

Kumar describes geodetically aligning calibrated reference 2D imagery with an 2D input image.

In claim 3, a portion of each video object in a particular frame is projected to intersect the projection of the video object in an adjacent frame to provide continuous silhouettes of the video object according to the time *t* coordinates. Kumar constructs mosaics from frames of a video by "aligning successive video frames using low order parametric transformations such as translation, affine and projective transformations." In the section cited by the Examiner, Kumar simply determines which pixels in a frame represent something that wasn't included in a previous frame, e.g., when a camera pans across a scene. The applicants request the Examiner specifically point out where Kumar projects "a portion of each video object in a particular frame," or "provide continuous silhouettes of the video object according to the time *t* coordinates." Kumar never determines video objects, or

silhouettes of video objects. Kumar adds new pixels from a next frame to a previous frame.

Claim 4 is rejected under 35 U.S.C. 103(a) as being unpatentable over Wu, Frisken, Kim, and Kumar in further view of Clatanoff, et al., (U.S. Patent 5,592,231 – "Clatanoff").

Clatanoff describes an interlaced-to-progressive-scan conversion process using motion-compensated interpolation. He describes performing motion detection with the use of median-filtered inter-frame difference signals and uses a fast median filtering procedure.

Median filters are known, but Clatanoff cannot cure the defects of Wu, Frisken, Kim or Kumar, see above. Clatanoff is silent as to growing, comparing and combining video volumes, projecting a portion of each video object in a particular frame to intersect the projection of the video object in an adjacent frame, and providing continuous silhouettes of the video object according to the time *t* coordinates as claimed.

Claim 5 is rejected under 35 U.S.C. 103(a) as being unpatentable over Wu and Frisken in further view if Kim (U.S. 5,978,031 – "Kim '031") and Gilhuijs, et el., (U.S. 6,112,112 – "Gilhuis").

In claim 5 the video is partitioned into a plurality of identically sized volumes and the pixel *at the center* of each volume is selected as the marker pixel. Kim '031 describes video object planes (VOP). A VOP can never be a volume because a

VOP is two dimensional and a volume is three dimensional. The two-dimensional VOP is described in Kim '031 at col. 1, lines 45-53 as follows:

> "A VOP can be referred to as an object and represented by means of a bounding [2D] ***rectangle*** whose ***width and height*** may be the smallest multiples of ***16 pixels surrounding each object*** so that the encoder processes the input ***video image*** on a VOP-by-VOP basis, i.e., an object-by-object basis." (*emphasis added*)

Therefore, Kim '031 can never be used to make three-dimensional volumes obvious. The Examiner's comment that partitioning the video into identically sized volumes facilitates block-based processing is non sequitor.

Gilhuijs describes a voxel-based method for assessing breast tumors using magnetic resonance imaging. A voxel is the 3D equivalent of a 2D pixel. A voxel is not a pixel. The coordinates of pixels are 2D in space and 1D in time, (x, y, t). The voxel coordinates of Gilhuijs are 3D in space only. See, for example, col. 4, lines 45-51, "Each volume consists of 64 slices of 256.times.256 pixels. The pixel size is 1.25x1.25mm^2, and the typical slice thickness is 2.0 mm." The claimed marker pixels are located at (x, y, t). Gilhuijs seed voxels are located at (x, y, z). Spatial coordinates can never make obvious spatio-temporal coordinates according to the invention.

Claims 8 and 9 are rejected under 35 U.S.C. 103(a) as being unpatentable over Wu and Frisken, in further view of Rosenfeld, et al., (Digital Picture Processing, Vol. 2, 1982 – "Rosenfeld").

Rosenfeld describes two-dimensional region merging at pages 141-142, and Figure 38. As stated above, regions are two-dimensional. Volumes are three-dimensional. The invention merges volumes less than minimum size with adjacent volumes in claim 8, and in claim 9 the minimum size is less than 0.001 of the volume representing the video. 2D region merging can never make obvious 3D volume merging.

Claim 10 is rejected under 35 U.S.C. 103(a) as being unpatentable over Wu, Frisken and Rosenfeld, in further view of McGee, et al (U.S. 6,496,228 "McGee") and Jewel (U.S. 5,852,683).

In claim 10 the volumes are sorted in an increasing order to size and processed in the increasing order by including each pixel of the volume in a closest volume until all volumes less than the minimum size are processed. McGee merges 2D regions at col. 13, lines 24-27. As stated above, two-dimensional region merging can never make obvious three-dimensional volume merging. Jewel merges two 2D images into a single 2D image. Jewel never processes 3D volumes. Jewel processes image regions.

Claim 11 is rejected under 35 U.S.C. 103(a) as being unpatentable over Wu and Frisken, in further view of Rudin, et al., (U.S. 6,452,637 – "Rudin").

Rudin also merges 2D regions, not volumes. Rudin can never make the invention obvious.

All rejections have been complied with, and applicant respectfully submits that the application is now in condition for allowance. The applicant urges the Examiner to contact the applicant's attorney at phone and address indicated below if assistance is required to move the present application to allowance. Please charge any shortages in fees in connection with this filing to Deposit Account 50-0749.

Respectfully submitted,

By: ______________________

Andrew J. Curtin
Reg. No. 48,485
Attorney for Assignee

Mitsubishi Electric
Research Laboratory, Inc.
201 Broadway
Cambridge MA, 02139
(617) 621-7539

Inventor(s): Porikli, et al. Attorney Docket: MH-5066

Filed: April 4, 2001 S/N: 09/826,333 Substitute Figure 1

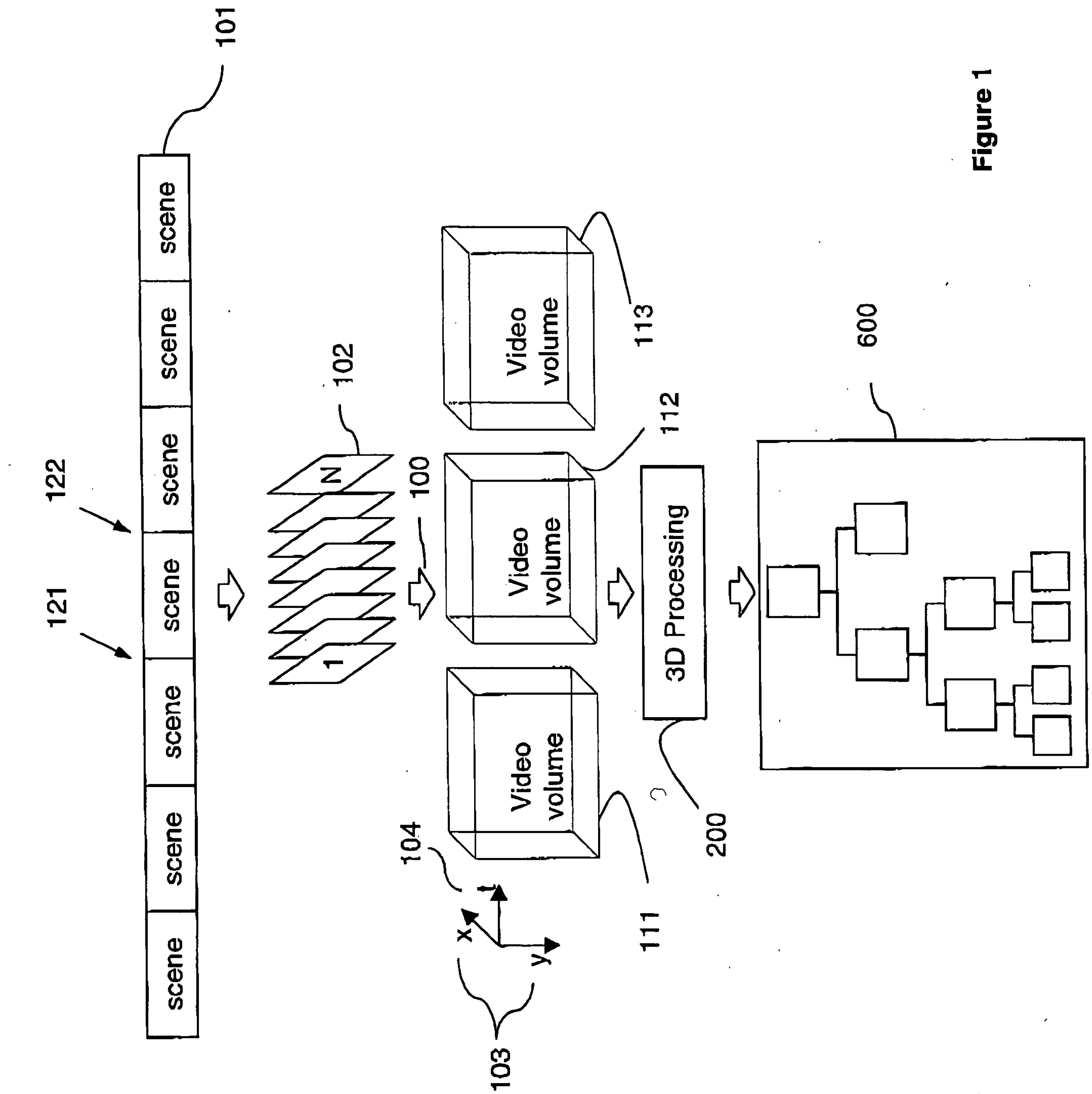


Figure 1